978-0-12-384719-5

Cladistics

Ian J. Kitching Peter L. Forey David M. Williams
~~The~ Natural History Museum, London, UK

Abstract

Cladistics is a class of methods of biological classification that groups taxa hierarchically into discrete sets and subsets. This article presents the principles and concepts of cladistics and describes the principal analytical methods. The operations by which observations of organisms are coded for analysis are explained, followed by the methods for reconstructing the hierarchical relationships among taxa (usually expressed as branching diagrams termed cladograms). Statistics and principles for determining the degree of fit between data and cladograms are discussed, which permit choices to be made among competing cladograms.

Keywords

Cladistics
Homology
Parsimony
Phylogenetic trees

Glossary

Apomorphy
: A derived character or character state; if two or more taxa share apomorphies, these are referred to as synapomorphies.

Clade
: Group of taxa diagnosed as monophyletic by the discovery of homologies (or synapomorphies).

Cladogram
: Branching diagram specifying hierarchical relationships among taxa.

Cladogram support
: Tests that permit some evaluation of how well data fit a cladogram.

Consensus cladogram (tree)
: Branching diagram that summarizes the common branching patterns from two or more cladograms.

Homology
: Two characters passing the similarity, conjunction, and congruence tests are termed homologous; in cladistics, homology is synonymous with synapomorphy.

Homoplasy
: A character or character state acquired by parallel or convergent evolution that bears resemblance to a character in a different group.

Monophyly
: Relationship between taxa united by a synapomorphy.

Optimization
: Procedure for reconstructing the most parsimonious sequence of character change on a cladogram.

Parsimony
: General scientific principle that given alternative explanations or hypotheses for a set of observations or data, the most corroborated is that requiring the fewest ad hoc (ancillary or additional) hypotheses.

Plesiomorphy
: An apomorphic character or character state that specifies a more inclusive group than that under consideration.

Relationships

The basic concept of cladistics is that genealogical connections among organisms are expressed in relative terms. Consider three taxa, A, B, C, whose genealogical relationships are as given in Figure 1a. Taxa B and C are more closely related to each other than either is to taxon A because they share a common ancestor, x

(which lived at time t_1), that is not shared with taxon A. Similarly, taxon A is more closely related to the group (B+C) because A, B, and C together share a unique common ancestor, y, that lived at an earlier time (t_0). In a real example (Figure 2), the human and turkey are considered to share a unique common ancestor (w) that lived at t_3. Similarly, the frog, turkey, and human are more closely related to each other than to either the perch or dogfish because these three taxa uniquely share a common ancestor x that lived at time t_2. The human and turkey are called sister-groups. Likewise, in this example, the frog is the sister-group of (human+turkey). The aim of cladistic analysis is to ~~infer~~ discover the sister-group hierarchy of life-forms by examination and analysis of their characters and to express the results as branching diagrams. These diagrams are called "cladograms" because they identify a hierarchical arrangement of taxa based on homologies termed "clades."

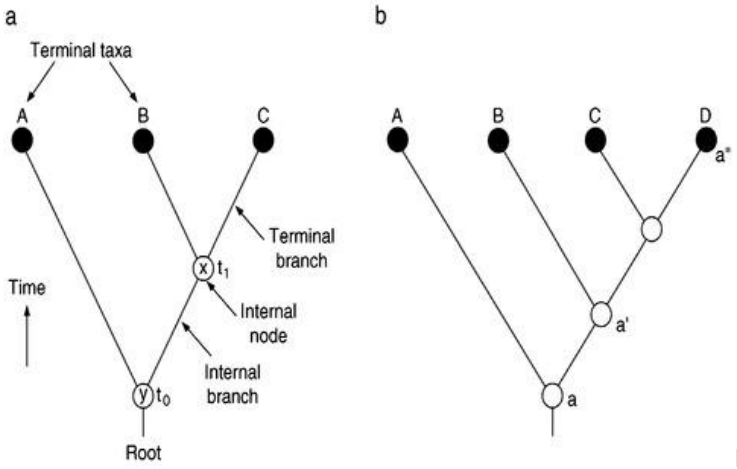

Figure 1. (a) The phylogenetic tree and the cladistic meaning of relationship. Taxa B and C are considered more closely related to each other than either is to taxon A because they share a unique common ancestor (x) that is not shared by taxon A. (b) Plesiomorphy and apomorphy are relative terms. On this phylogenetic tree, a′ is apomorphic with respect to a, but plesiomorphic with respect to a″ .

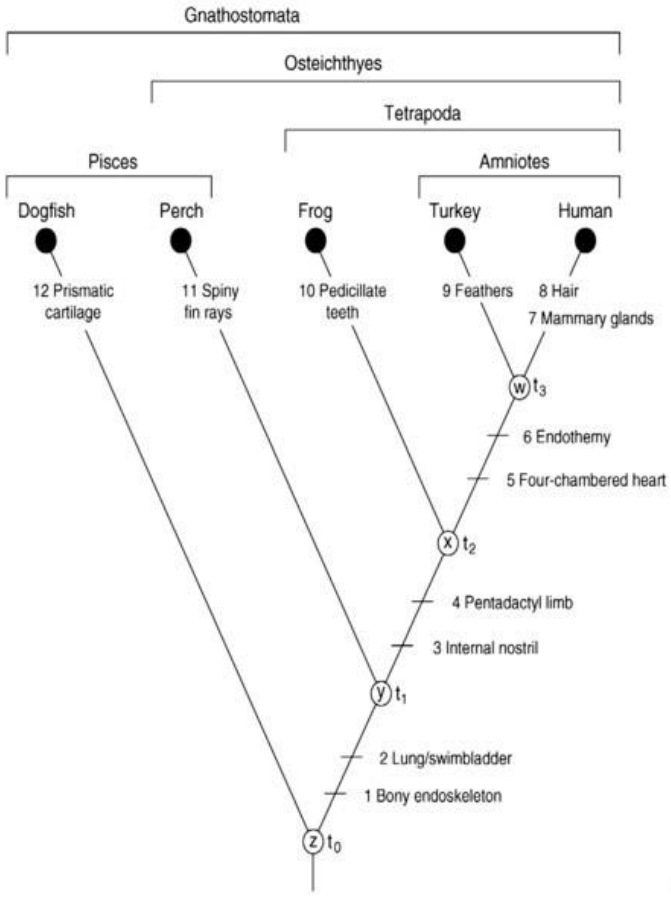

Figure 2. A phylogenetic tree for five taxa of vertebrates. Three monophyletic groups are established using characters 1–6, while autapomorphic characters 7–12 diagnose the terminal taxa. The group Pisces is paraphyletic because one of its included members (the perch) is cladistically more closely related to Tetrapoda.

Types of Characters

Two types of characters are recognized based on where they occur on a cladogram (Figure 1b). The character that occurs in the ancestor is termed “plesiomorphic” (near to the ancestral morphology)

and the derived character is “apomorphic” (away from the ancestral morphology). Apomorphic and plesiomorphic are relative terms, that is, relative to a particular systematic problem. In Figure 1b, character a′ is apomorphic with respect to character a, but plesiomorphic with respect to character a″ .

Cladistic analysis proceeds by identifying shared apomorphic characters or “ synapomorphies.” In Figure 2, a four-chambered heart and endothermy are synapomorphies that suggest the human and the turkey share a unique common ancestor, w. The cladogram implies that these two characters arose in ancestor w and were then inherited by both the human and the turkey. Synapomorphies can ~~may~~ therefore be considered ~~as~~ evolutionary homologies. In contrast, the shared possession of internal nostrils and pentadactyl limbs by the human and turkey does not imply that they share a unique common ancestor because these attributes are also found in the frog. These shared primitive characters (or “symplesiomorphies”) are inherited from an ancestor more remote than the most recent common ancestor of the human and the turkey. They are thus irrelevant to the hypothesis of a relationship between the human and the turkey. However, with respect to the more inclusive three-taxon problem comprising the frog, turkey, and human, internal nostrils and pentadactyl limbs are relevant. At this level, they are synapomorphies suggesting that these three taxa form a group with a common ancestry at x.

Apomorphies occurring in only a single terminal taxon are termed “autapomorphies. ” In Figure 2, these are prismatic cartilage (dogfish), spiny fin rays (perch), pedicillate teeth (frog), feathers (turkey), and hair and mammary glands (human). However, if a terminal taxon is itself a group, then its autapomorphies are also synapomorphies of its component taxa.

Thus, as with the cladistic meaning of relationship, characters are also relative, depending on the systematic problem under consideration. Furthermore, it should be stressed that characters are derived from observations of the features occurring in organisms. These are the homologues. ~~and~~ The relationship implied by the homologues can be referred to as homology (Williams, 2004). ~~t~~The explication of the~~ir~~ hierarchical distribution of homologues need not imply a particular theory of evolution.

Parsimony

Relationships among three taxa (as in Figure 1a) can be resolved in three ways—A (B C), B (A C), and C (A B)—whereas for four taxa (as in Figure 1b) there are fifteen possible fully resolved cladograms. In cladistic analysis, parsimony is the universal criterion for selecting among alternative hypotheses of character distribution. Characters are fitted onto alternative topologies and the cladogram that accounts for the greatest number of characters in the simplest way is chosen as the best hypothesis of relationships.

Suppose six characters are distributed among four taxa as shown in the taxon/character matrix in Figure 3. Taxon A has none of the characters but the other three taxa each have a different complement. Characters 2 and 4 are autapomorphies because they are each present in only one of the taxa. They are uninformative for grouping taxa (they serve only to diagnose the terminal taxa). Characters 1, 3, 5, and 6 are potentially useful because they occur in more than one taxon. Given the three taxa that have potentially informative information, there are three ways in which these taxa can be arranged dichotomously (Figure 3b–d). If the characters are now placed onto each possible cladogram, according to the groups they specify, then three different results are obtained (Figure 3e–g). In Figure 3e, all characters except one appear only once. In this solution, character 6 must be assumed to appear twice, once in taxon B and once in taxon C, which are not sister-groups. In this example, the behavior of character 6 is homoplastic, that is, it occurs more than once on the cladogram and is said to be a homoplasy. In contrast, in the other two topologies (Figures 3f and 3g), we must assume that two or more characters appear more than once. Hence, the cladogram in Figure 3e accounts for the distribution of the characters in the most economical way and is thus the preferred solution.

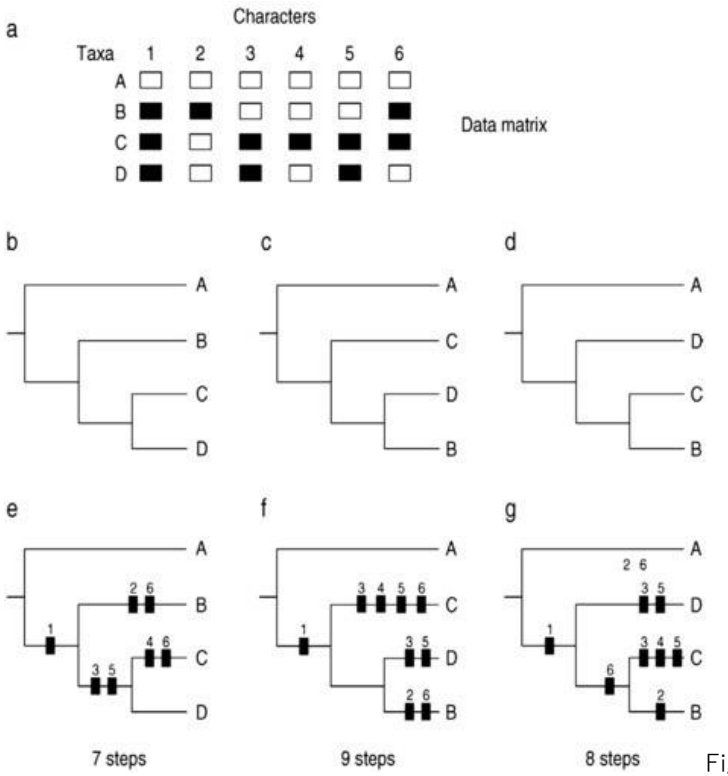

Figure 3. Parsimony. (a) A data matrix of six characters distributed among four taxa, A–D. Plesiomorphic states are shown by open boxes and apomorphic states by black boxes. Taxon A is totally plesiomorphic. (b–d) The three possible resolutions of taxa B–D relative to A. Placing the characters on these three topologies results in (e) being selected as the optimal cladogram while (f) and (g) are suboptimal.

The distribution of characters can also be regarded as the number of steps on a cladogram, which, in Figure 3, is the number of instances where a character is gained. In Figure 3e, this is seven, while the other cladograms (Figures 3f and 3g) are more costly, requiring nine and eight steps, respectively. The concept of steps is actually a little more subtle than the sum of character gains because a character may appear at one point on a cladogram and then disappear at another point. For example, another explanation for the distribution of character 6 in Figure 3e is to

assume that it is gained by the group (B+C+D) and then lost in taxon D. Each change, whether gain or loss, is considered a step. In this example, both accounts of character change demand two steps and therefore both hypotheses of character change are equally parsimonious. The sum of the number of steps on a cladogram is termed the length of the cladogram, irrespective of whether the changes are gains or losses. The most parsimonious solution is also known as the optimal cladogram and the other cladograms (i.e., those requiring more than the minimum number of steps to explain the character distributions) as suboptimal.

It is possible for a given set of characters to yield two or more equally most parsimonious cladograms. Then, we may prefer to accept one of the solutions based on other criteria, such as a closer agreement with the stratigraphic record or by differential weighting of one type of character change relative to another. Alternatively, we may simply accept that the conflict in the data is such that we cannot derive a unique most parsimonious solution. For certain purposes, we may choose to combine those components common to the different solutions to form a consensus cladogram.

Groups

Cladistics recognizes only monophyletic groups of organisms, which are those based on synapomorphies. Monophyletic groups are the only groups that can be circumscribed by objective boundaries, defined by characters. In evolutionary terms, monophyletic groups comprise the most recent common ancestor and all of its descendants. In Figure 2, Amniota, Tetrapoda, Osteichthyes, and Gnathostomata are all monophyletic. Two other types of “groups” are sometimes referred to but these are not groups in the same sense as monophyletic groups. Paraphyletic “ groups” are based on symplesiomorphy; in evolutionary terms, their members are linked by common ancestry but one or more of the descendants of the most recent common ancestor are excluded. In Figure 2, Pisces (fishes) is a paraphyletic assemblage. Many taxa traditionally regarded as ancestral, such as fishes, reptiles, and green algae, are paraphyletic. Polyphyletic “groups” are based on homoplasy, that is, characters that are considered convergently derived and that cannot be inferred to have been present in the most recent common ancestor of the included

taxa. In Figure 2, an assemblage comprising the dogfish and the turkey (perhaps based on the observation that both lay eggs surrounded by a shell, although no one would claim such a homology) would be a polyphyletic group.

Cladograms and Phylogenetic Trees

A cladogram is a diagram that summarizes a pattern of character distribution. Usually, a cladogram is drawn as a branching diagram (e.g., Figure 1). The nodes denote a hierarchy of synapomorphies but, on its own, there is no necessary implication of ancestry and descent. Cladograms may also be written in parenthetical notation or illustrated as a Venn diagram (Figure 4a), which conveys the same grouping information as a branching diagram. In contrast, phylogenetic trees include a time axis and embody concepts of ancestry and descent with modification. In phylogenetic trees, the nodes denote ancestors (known or hypothetical) and the branches imply character change. Several phylogenetic trees may be compatible with the pattern of character distribution implied by a cladogram (Figure 4b). Some of these trees allow the possibility that one or more taxa are ancestral to others. Only the phylogenetic tree that assumes all nodes represent hypothetical ancestors has the same topology as the cladogram. Thus, cladograms are more general than phylogenetic trees, which are precise statements about ancestry and descent.

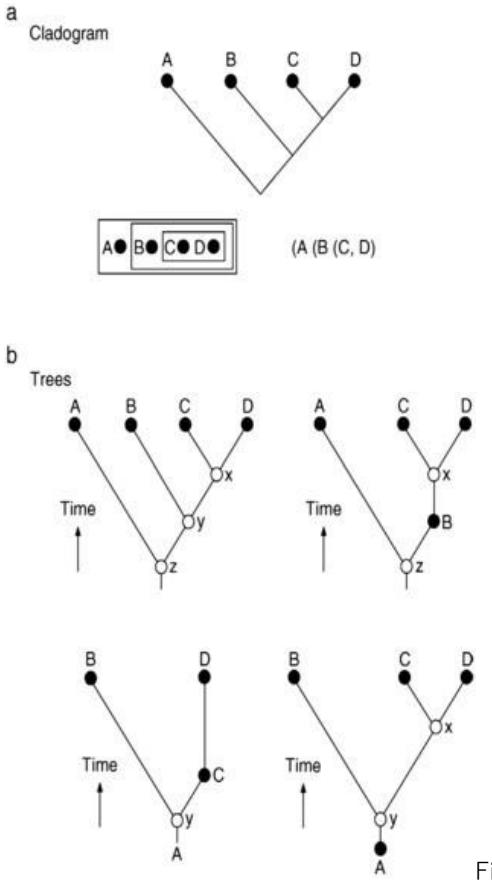

Figure 4. Cladograms and trees. (a) A cladogram simply shows the hierarchical distribution of characters. It has no time axis and may be drawn as a branching diagram, a Venn diagram, or in parenthetical notation. Given this cladogram, several phylogenetic trees may be inferred from the same data. Some of these are shown in (b). All except that on the top left invoke the concept of one of the taxa being ancestral. The tree at the top left assumes only hypothetical ancestors and is the only one that is synonymous with the cladogram.

Characters and Coding

Opinions differ over the nature and discovery of taxonomic characters. One view holds that characters are properties of organisms that provide quantifiable variation. Alternatively, characters may be viewed as theories concerning two (or more) attributes, which may look different but are nevertheless considered the same. This latter view is embraced within a general understanding of homology, such that characters may be understood as homologues~~in the same manner as homology~~ and the implied relationship as homology. There is a lack of agreement over what indicates the discovery of a character. However, all definitions of homology suggest that it concerns features that are similar in different taxa sharing a recent common ancestor. Such definitions satisfy as explanations but do not aid discovery.

Homology

Within cladistics, various tests have been proposed to assist in establishing homology. One view of characters is that they are identical in meaning and discovery to homology, and homology may be conceived as a series of three tests that apply to methods of comparison.

The similarity “test” suggests that without evidence to allow direct comparison of one feature with another, there would be no proposition of homology and, consequently, no concept of a character. This “test” is not exact and cannot be taken to imply “identity.” For example, comparisons may consider the detailed similarity of any two stamens or the inferred similarity of mammalian stapes with gnathostome hyoid arches.

The conjunction test suggests that two features that co-occur in the same organism cannot be considered homologous. A familiar example, albeit contrived, is an angel with both wings and forearms. The two kinds of limbs in the same individual cannot be considered homologous. Many comparisons fail this test and are often associated with “homomony” or serial homology (e.g., the individual vertebrae of a single vertebral column or the abdominal appendages of arthropods).

The congruence test is considered the most exacting and refers to the support afforded to one homology by others. In other words, homologies are considered to have passed the test if there are other homologies that specify the same

~~taxon~~group. Congruence is actually an analytical procedure and is usually considered in terms of parsimony. However, it also points to another property of homology, that is, homology can never be proven. As data ~~are~~ accumulate~~d~~, previously supported homologies may be overturned and new theories of homology established in their place.

The tests of homology may be more understandable if they are applied to homologues (the parts) instead of homology (the relationship).

Character Recognition

It is generally agreed that characters, however conceived, are based on observations. Stated simply, a feature (e.g., stamens, shoulder girdles, wings) is observed in a particular specimen and directly translated into the character. This approach may initially seem useful and would eventually lead to enumeration of all features of the specimen. However, the final list would not consist of "characters" but would be an inventory of "features," each being a descriptive element of the specimen and implying that such descriptions apply to all specimens of the same taxon. Each descriptive element contains a notion of theory. These elements would be the homologues.

Suppose the specimen examined is a rat. Initially, it would be straightforward to describe: head, body, limbs, tail, and so on. More detailed examination would reveal, for example, a vertebral column. We identify the vertebral column by drawing on knowledge of previous studies of rat anatomy and are able to confirm its detailed similarity to other vertebral columns. In so doing, we assimilate what is already known of vertebrates: they are animals with a vertebral column. If this process were performed for all features, then it would seem that all attributes of this single rat could uncover its place in the hierarchy of life, at every inclusive level. In this sense, taxonomic characters are very much like homologues ~~homologies~~ when interpreted in the light of taxa: features that (potentially) specify a particular taxon. A vertebral column does not tell us it is a rat; it tells us that it is a vertebrate. In this sense, the vertebral column is a *feature* of any particular rat but only a *character* of vertebrates. This distinction identifies the general task of systematics: to identify the level at which various attributes (homologues) ~~are homologies (characters)~~ diagnos~~e~~ing taxa (homology).

Kinds of Characters

Characters are often thought of as comprising different types. Some refer to different numbers of a feature, and others refer to differences in structure. For example, variation in stamen structure in angiosperms encompasses both different forms of anthers and filaments and differences in their numbers. This exemplifies the distinction between quantitative and qualitative characters, the former usually being counts ("meristic characters") or measurements ("biometric characters"), the latter relating to structural differences. Quantitative characters may be problematic for cladistics insofar as it can be difficult to render measurements and counts as meaningful homology statements. This is not to say that such characters are not useful, for they can serve to identify particular specimens. However, their use may be limited because they are not always amenable to cladistic analysis. An alternative view separates characters into neomorphic (presence/absence) and transformational (variations on a present structure) (Sereno, 2007; see also Cladistic Analysis: Character Coding).

Characters as Phylogenetic Evidence

Although structural evidence is sought for cladistic purposes, the observed features themselves are not necessarily the characters. For example, some organisms have fins, others have arms, and yet others have wings. Studies of fins, arms, and wings show that they have certain parts in common as well as certain parts that are unique. These common properties might suggest an initial proposition that fins, arms, and wings are all examples of a single character, in this case "paired appendages." However, further details are needed to confirm this hypothesis. Suppose the "wings" examined were from an insect such as a housefly. It might still seem reasonable to call them "paired appendages," but in this example there is nothing in the detailed anatomy to suggest that housefly wings and mammalian arms (or even bird wings) are in any way "the same." One conclusion is that the term "wings" is ambiguous when describing attributes of organisms but not when describing the function of these attributes (here, flight). Wings may indeed be considered as a part of an organism but flight is their assigned function (usually). A

more reasonable conclusion is that “wings” is not a character at all but a functional attribute. Thus, the wing of a bird is better considered as a modified “paired appendage,” modified for flight. The problem, however, is not simply semantic. If the comparison was made between a bird and a bat, then detailed anatomy does suggest that both are indeed “paired appendages” and that both are modified for flight. Our current understanding of vertebrates suggests that birds and bats do not form a monophyletic group. Hence, bird wings and bats wings would be considered two characters rather than one when interpreted on the cladogram (Figure 5). Their “sameness” is captured as “forelimbs,” their differences as wings of birds and wings of bats.

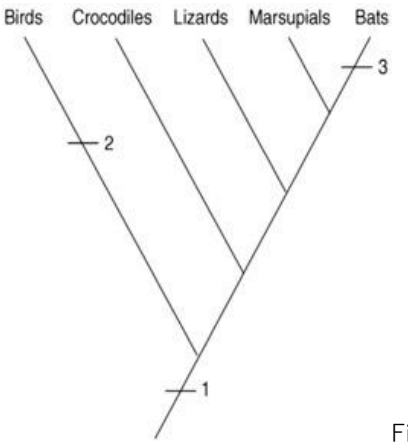

Figure 5. Relationships of vertebrates; 1 ≡ “forelimbs”; 2 ≡ “wings of birds” 3 ≡ “wings of bats.”

Consideration of the wings of birds, the forelimbs of tetrapods, and the fins of fishes together identifies a well-defined character (“paired appendages”) with various manifestations. These manifestations might suggest particular taxonomic groups. There might be a taxon with “fins,” a taxon with “arms,” and a taxon with “wings.” This, of course, was precisely the situation for many years: fishes (Pisces) have fins, tetrapods have arms, and birds (Aves) have wings (Table 1). Derived from these observations is the implication that fins, wings, and arms are in some way connected, other than by being “paired appendages”:

fins – arms – wings

Furthermore, that connection might be viewed in an evolutionary context, such that it represents the transformation of one manifestation (e.g., fins) into another (e.g., arms):

fins → arms → wings

In more general terms, features could be represented as modifications of other features, such that they are hierarchically related:

Fins
Standard fins (fins)
  Modified fins (arms)
    Modified arms (wings)

Table 1.

Taxon	Character
Pisces	Fins
Tetrapods	Arms
Birds	Wings

The features are no longer structured as a series of alternatives, as in the first example (fins–arms–wings), but now specify a nested set of relationships. The proposition is that wings are really kinds of arms, and that arms are really kinds of fins, and fins represent the entire set of animals with “paired appendages.” It is possible to interpret all of these taxa (fishes, tetrapods, and birds) as having fins. Hence “fins” is not a character of fishes but rather of gnathostomes (in this example, fishes+mammals+birds). Consequently, both the character “fins” and the taxon Pisces disappear. This view confirms the notion that characters are hypotheses drawn from observations rather than simply the observations

themselves. Such hypotheses are identical to those made for general statements of homology. Superficially, the relationship between fins, arms, and wings may be considered identical to “fins→arms→wings,” as was implied by Hennig (1966) in his concept of “transformation series.” However, it is possible to view characters as more general, specifying particular relationships in terms of a definitive statement connecting to a taxon.

Character Coding

One significant outcome of theories relating to characters is how they might be represented numerically for cladistic analysis. For example, one might code each “character” (fins, wings, arms) separately (Table 2, characters 2–4). This scheme reflects the “uniqueness” of each attribute but contains no information relevant to recognizing that the three observed forms are connected as “paired appendages.” This approach is referred to as “absence/presence binary coding,” because a positive value (usually 1) is assigned to the presence of a feature and a negative value (usually 0) is assigned to the absence of the feature. Alternatively, one might represent the same series of observations in a single column to signify their connection (as “paired appendages”), then assign each unique feature a separate value (Table 2, character 1). This is “multistate coding” and considers the character to be composed of discrete states that bear some (usually unspecified) relationship to one another. Hence different values appear in the same column and are treated as dependent on the other values. This might not be seen as completely sufficient, as additional information would be needed to specify the exact nature of the connection. For instance, one might wish to specify that the “characters” are connected but that the nature of that connection is unknown. Choices of this nature relate to character optimization (see Cladistic Analysis: Optimization).

Table 2.

Taxon	Characters			
	1	2	3	4
Fins	1	1	0	0
Wings	2	0	1	0

Taxon	Characters			
	1	2	3	4
Arms	3	0	0	1

To summarize, characters have their origin, but not their identity, in observations. Characters are what lead us to suspect that taxa exist (vertebral column ≡ vertebrates ≡ taxon Vertebrata) and hence are identical to conjectures of homology derived from empirical investigation of specimens (of their homologues). Homology is the relation that specifies taxa and that implies an intimate relationship between characters and taxa. Both are the results of analyses and are discovered by our investigation of features.

Cladistic Analysis

Cladogram Construction

The original method of cladogram construction was proposed by

Hennig (1950, 1966) x
Hennig, 1950
Hennig, 1966

Hennig (1950, 1966) and is thus known as Hennigian argumentation. In this approach, characters are first polarized into plesiomorphic and apomorphic states. The groups thus diagnosed by synapomorphies are then organized manually into a cladogram. However, this procedure can only find the most parsimonious cladograms when the data are free or nearly free of homoplasy (i.e., the fit of data to most parsimonious cladogram is perfect or nearly so). For larger and more complex data sets, computerized algorithmic methods become a necessity.
There are two main computerized approaches to cladogram construction. Exact methods guarantee to find the most parsimonious cladograms. The simplest exact method is “exhaustive search.” First, three taxa are chosen and connected to form the only possible unrooted, fully resolved cladogram for these taxa. Then, a fourth taxon is selected and added to each branch to yield the three possible fully resolved, partial, unrooted cladograms for four taxa. A fifth taxon is then selected and added to each of the five branches on these three partial cladograms to yield

the fifteen possible fully resolved unrooted topologies for five taxa. This process is continued, following every possible path of taxon addition, until all taxa have been added and all possible fully resolved cladograms have been found. The lengths of these cladograms are then calculated and the shortest is chosen as the optimal solution(s). However, as the number of taxa increases, the number of cladograms to be examined rises exponentially and the time required for exhaustive search soon becomes unreasonable.

One exact method that does not require every possible cladogram to be evaluated is "branch-and-bound" analysis. In this approach, a preliminary cladogram is constructed and its length is set as the upper bound for subsequent searches. A procedure similar to an exhaustive search is then undertaken but at each step the length of the partial cladogram is recorded. Whenever this length exceeds the current upper bound, that partial cladogram is rejected (and so, consequently, are those complete topologies that would be derived from it by adding the remaining taxa). By this means, the number of topologies to be examined is reduced. Once all taxa have been added, the length of the complete cladogram is examined and if it is equal to the upper bound, then that topology is retained as a most parsimonious cladogram. However, should this cladogram be shorter than the current upper bound, then its length is substituted as a new upper bound. This important procedure allows subsequent partial cladograms to be rejected quickly and thus speed up analysis. This process continues until all possible paths have been examined, whence the set of optimal cladograms will have been found.

For large data sets (more than 30 taxa), even branch-and-bound analysis can be too time-consuming. In this case, approximate or "heuristic" methods are used. These approaches examine only a subset of all possible topologies and thus are not guaranteed to find the most parsimonious cladogram(s). However, they are faster than exact methods for large numbers of taxa and thus certainty of finding the optimal cladogram(s) is sacrificed for decreased computational time.

Heuristic analysis comprises two stages. In the initial building phase, a cladogram is constructed using a process of "stepwise addition." The order in which taxa are added is termed the "addition sequence" and there are various ways in which taxa may be added. Once a complete cladogram has been constructed, attempts can be

made to improve upon it by performing a series of rearrangements called “branch-swapping.” The cladogram is cut into two or more partial cladograms, which are then recombined in order to try to find new, shorter topologies. Other methods, such as the Parsimony Ratchet, use differential weighting to search more efficiently for the shortest cladograms, and algorithms such as Tree Fusing and Tree Drifting offer yet further improvements. The efficiency of ~~current branch-swapping~~these algorithms in finding most parsimonious cladograms is ~~very~~ extremely high, but it is always possible that they can become trapped in a local optimum. Thus, one should always be aware with heuristic analyses that shorter topologies than those reported may exist.

Character Polarization and Cladogram Rooting

Manually implemented cladistic methods, such as Hennigian argumentation, require synapomorphies to be identified in advance of cladogram construction. The process through which plesiomorphic and apomorphic characters are distinguished is termed “character polarization.” Numerous criteria for polarizing characters have been proposed, but only two are now considered valid.

The first criterion, called “outgroup comparison,” was classified by Nelson (1973) as an “indirect” method, because it draws upon evidence from a source (the “outgroup ”) that is external to the taxa under investigation (the “ingroup”). In its most basic form, polarization using outgroup comparison can be defined as follows: “For a given character with two or more states within a group, the state occurring in related groups is assumed to be the plesiomorphic state” (Watrous and Wheeler, 1981: 5). This definition is adequate when all outgroup taxa share the same state, but it is insufficient if the outgroup taxa are heterogeneous. Maddison *et al.* (1984) further noted that it is inappropriate to estimate the state in the most recent common ancestor of the ingroup (the “ingroup node”). Rather, it is the state at the next most distal node, linking the ingroup to the first outgroup (the “outgroup node”), that should be estimated if the solution is to be globally optimal, and they described an algorithmic approach to such reconstruction.

In contrast, Nelson (1973) classified the “ontogenetic criterion” as a “direct” method, because its implementation relies on evidence from the ingroup taxa alone.

It is defined as follows: “Given an ontogenetic character transformation from a character observed to be more general to a character observed to be less general, the more general character is primitive (plesiomorphic) and the less general character derived (apomorphic)” (Nelson, 1978: 327). For example, the embryos of both sharks and frogs have cartilaginous skeletons. However, this condition persists into the adult shark, but in frogs the cartilage is replaced by bone during ontogeny. In this example, a bony skeleton is observed to be less general (occurring only in the frog) and is thus interpreted as apomorphic. In this context, “more general” is defined as that occurring earlier in ontogeny. As such, the more general character is not simply the more common (although this may often be the case) and the ontogenetic criterion does not equate with commonality. What is important is that the less general character is nested within the observed distribution of the more general character. This requirement is violated by ontogenetic sequences that are secondarily simplified through paedomorphosis or neoteny. Such an ontogeny cannot be interpreted as proceeding from the more general to the less general, and Nelson's criterion will not allow us to distinguish a secondarily reduced ontogeny from the plesiomorphic sequence. Then, we rely on congruence with other characters to make the distinction.

Most recent cladistic studies do not actually include a priori polarization of characters, but undertake “simultaneous, unconstrained analysis”: “simultaneous” because both ingroup and outgroup taxa are analyzed together, and “unconstrained ” because outgroup taxon relationships are unspecified before analysis. Cladograms are then rooted between the outgroup node and the remaining outgroup taxa, at which point character polarities are established.

Optimization

Optimization is the process of determining the sequence of character state changes on a cladogram in order to test hypotheses of transformation. If the data include characters coded as multistate, then these may be interpreted according to many different optimality criteria, of which the best known are Wagner and Fitch optimization.

Wagner optimization (Figure 6a–b) is used for “ordered” or “additive” multistate characters, in which transformations between successive states are considered as incremental. Thus, the changes 0↔1 and 1↔2 each “cost” the same number of steps (usually one), whereas the change 0↔2 is considered to “pass through” state 1 and thus costs two steps. Costs are symmetrical, so that the changes 0→1 and 1→0 both constitute a single step (termed “free reversibility”).

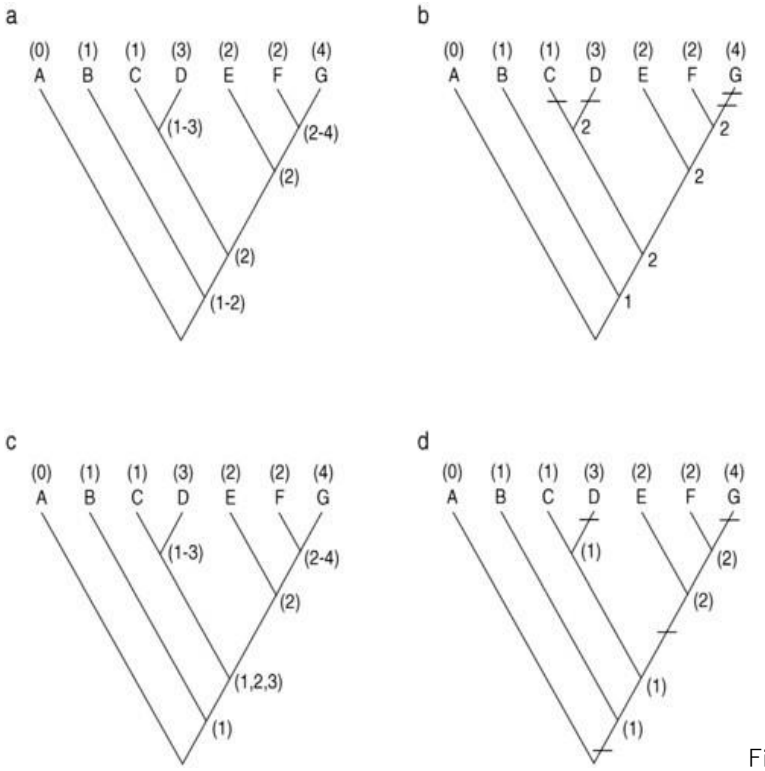

Figure 6. Character optimization, (a) Wagner optimization (ordered multistate characters). State sets are assigned to internal nodes on the first pass from the terminal taxa toward the root. (b) States are assigned to the internal nodes on the second pass from the root to the terminal taxa. (c) Fitch optimization (unordered multistate characters). State sets are assigned to internal nodes on the first pass from the

terminal taxa toward the root, (d) States are assigned to the internal nodes on the second pass from the root to the terminal taxa.

Wagner optimization is implemented in two stages. First, the minimum number of steps for a character is determined. On a pass down the cladogram, from the most distal taxa toward the root, each of the internal nodes is assigned a “state set” (Figure 6a), which is defined as the intersection of the two derivative state sets. If the intersection is empty, then the smallest closed set that includes an element of each derivative state set is assigned. For example, the intersection of the state sets of taxon F (2) and taxon G (4) is empty. Thus, the smallest closed set, (2–4), is assigned to the node linking these two taxa. In contrast, the intersection between this state set and that of taxon E is not empty. Both contain state 2, and this value is assigned to the node joining taxa E, F, and G. Unambiguous states are then assigned to internal nodes by a second pass, up the cladogram (Figure 6b), to produce a “most parsimonious reconstruction.” If the state set of a node contains more than one value, then the node is assigned the value that is closest to that of the node of which it is a derivative. Thus, the node joining taxa F and G is assigned state 2, because this is the value of the next most distal node. The most parsimonious reconstruction has six steps because the character is ordered. As a result, the change from state 2 to state 4 along the branch leading to taxon G counts as two steps.

Fitch optimization (Figure 6c–d) is used for “unordered” or “non-additive” multistate characters in which transformations between any two states are considered equal. Thus, the changes 0↔1, 1↔2, and 0↔2 all cost a single step. As with Wagner optimization, costs under Fitch optimization are freely reversible.

Fitch optimization follows similar procedures to Wagner optimization but with two differences. First, the state set assigned to an internal node is the union of the two derivative state sets (Figure 6c). Second, when determining the most parsimonious reconstruction, a node with an ambiguous state set is assigned the value of the next most distal node if that value is an element of the ambiguous state set. Otherwise, an element is selected arbitrarily. This most parsimonious reconstruction (Figure 6d) has four steps because the character is unordered. As a result, the change from

state 1 to 3 along the branch leading to taxon D and the change from 2 to 4 along the branch leading to taxon G each count as a single step.

These procedures do not necessarily yield a unique most parsimonious reconstruction. For example, it is equally parsimonious to optimize the ordered character using the nodal state reconstructions shown in Figure 6d for the unordered character. State 1 is assigned to the node joining taxa C–G, and is followed by a two-step change, 1→3, on the branch leading to taxon D and a one-step change, 1→2, on the branch joining taxa E–G. The steps on the branches joining taxa B–G and leading to taxon C would be lost, thus maintaining the most parsimonious length of six steps. This type of optimization, in which changes are placed onto the cladogram as far from the root as possible, is called “delayed” or “ slow” transformation. In contrast, “accelerated” or “fast” transformation places changes onto the cladogram as close to the root as possible, as in Figure 6b. When alternative most parsimonious reconstructions are possible, character optimization may lead to “spurious resolution,” in which groups appear to be resolved but have no unambiguous support in the data. Such groups are not strong hypotheses of relationship, and Nixon and Carpenter (1996) suggested that they should be eliminated wherever possible (without violating the minimum length requirement). Those cladograms that remain, which are both of minimum length and have all groups supported unambiguously by data, are termed “strictly supported cladograms” and are the preferred topologies.

Cladogram Evaluation

Character Fit

The preferred solution to a cladistic analysis is the most parsimonious cladogram because this represents the simplest explanation of the data with the number of *ad hoc* hypotheses of homoplasy kept to a minimum (see Figure 3e). Consequently, the most basic measure for assessing the fit of data to a cladogram is cladogram length. The most parsimonious cladogram has best fit because it is the shortest; longer cladograms have poorer fit. However, the length of the most parsimonious cladogram is partly dependent on the absolute size of the data set from which it is derived. Larger data sets will necessarily yield longer cladograms than smaller data

sets. A binary character will display perfect fit when it is placed on a cladogram with a single step. Homoplasy is manifest as an increase in the number of steps. The amount of homoplasy implied by a character on a cladogram is measured by the " consistency index" (ci), which is the ratio of the minimum number of steps required by the character (m=1 for a binary character) to the observed number (s). In Figure 3e, character 5 occurs only once and hence its ci=1 (m/s=1/1), whereas character 6 shows two steps and thus its ~~ci~~ci=0.5 (m/s=1/2). The amount of homoplasy implied by the whole data set can be measured using the "ensemble consistency index" (CI), which is the ratio of the minimum number of steps implied by all characters (M) to the length of a cladogram (S). For the cladogram in Figure 3e, the CI=0.86 (M/S=6/7).

There are three perceived problems with the consistency index as a measure of homoplasy. First, although uninformative characters do not add any grouping information to a cladogram, they will inflate its CI. However, this is of significance only when different data sets are being compared. Second, CI can never attain a zero value. A data set in which all possible informative characters occur in equal numbers (an "undecisive" matrix) provides no evidence for preferring one cladogram to any other. Nevertheless, these cladograms will all have positive, non-zero CI values. Third, it has been observed empirically that CI decreases as the number of taxa increases, irrespective of change in the information content of the data. However, this is a recognized and expected property of the CI.

Although the consistency index is useful as a measure of the amount of homoplasy in a character or data set, it is indifferent to the pattern of fit on a cladogram. A binary character that occurs on two separate terminal branches will have the same ci value (0.5) as one that supports two separate groups of taxa. However, in the former case, the character contains no grouping information, while in the latter it is a synapomorphy (albeit homoplastic) for two groups of taxa. The amount of similarity in a character that is interpreted as synapomorphy is measured by the " retention index" (ri). This is defined as $(g-s)/(g-m)$, where s and m are the same variables as for ci, and g is the maximum number of steps that a character can show on any cladogram. For character 5 in Figure 3e, the minimum and observed number of steps is one, and the maximum number is two. Hence its ri=1 ($(g-s)/(g$

$-m$)=(2 − 1)/(2 − 1)) and all similarity is interpreted as synapomorphy. In contrast, for character 6, the minimum number of steps is one, and the observed and maximum number is two. Hence its ri=0 (($g-s$)/($g-m$) =(2 − 2)/(2 − 1)) and none of the similarity is interpreted as synapomorphy. The method can be extended to the whole data set as the “ensemble retention index” (RI), which uses the summed values of g, s, and m (G, S, and M, respectively). For the cladogram in Figure 3e, the RI=0.67 (($G-S$)/($G-M$)=(9 − 7)/(9 − 6)).

Character Weighting

The application of differential weights to characters has a long history in systematics. Methods of weighting can be divided into a priori and a posteriori procedures, depending on whether they are applied before or after cladogram construction.

A priori approaches to character weighting generally invoke beliefs that some characters are more important than others or use a particular model of evolution or character change, under which certain types of transformation are considered more or less likely than others. For example, it is common when analyzing nucleotide sequence data to downweight transition substitutions relative to transversions. Alternatively, changes in third codon positions may be disregarded (i.e., accorded zero weight) because they are considered much more likely than changes in first or second positions as a result of the redundancy of the genetic code. Numerous other models have been proposed and they are particularly frequent in the field of molecular systematics. However, such weighting schemes are justifiable only insofar as their underlying model is justifiable.

A posteriori weighting schemes are based on “cladistic consistency” (i.e., the fit of characters to a cladogram) and characters with greater fit are accorded greater weight. One indication of a character's fit is the amount of homoplasy it shows on a cladogram. However, homoplasy does not imply that all similarity is uninformative and the proportion of similarity interpreted as synapomorphy also needs to be taken into account. Hence, both the consistency index (homoplasy) and the retention index (synapomorphy) can be used to estimate character weights. Farris (1989) suggested using the product of these two measures, the “rescaled consistency

index" (rc). By combining the ci and ri in this way, characters in which none of the similarity is synapomorphic (ri=0) receive zero weight, irrespective of their level of homoplasy. All other characters, which contain some amount of grouping information (ri > 0), are differentially weighted according to their level of homoplasy. Using this approach, in Figure 3e character 5 would receive a weight of 1 (ci=ri=1), but character 6 would receive a weight of 0 (ci=0.5, ri=0). These weights are applied in a new analysis and the most parsimonious cladogram(s) obtained are used to estimate a new set of weights. This procedure is repeated until a stable set of both weights and most parsimonious cladograms is achieved; hence the name "successive approximations character weighting."

The level of homoplasy of a character may also be viewed as the number of extra steps required to fit it to a cladogram. If all extra steps in all characters are considered equal, then a linear fitting function is being applied to their relative cladistic consistency. For example, in Figure 3e, the single step of character 5 is considered equal to each and either of the two steps of character 6. However, intuitively, we might consider that characters showing fewer extra steps are "better" than those showing more. The former can be assigned higher weights than the latter using a concave fitting function of relative cladistic consistency. This approach was implemented by Goloboff (1993) as "implied weighting," in which the weight (W) accorded to a character is calculated as $W=K/(K+ESi)$. ESi is the number of extra steps shown by the character and K is the "constant of concavity." The value of K can be varied to weight more or less strongly against those characters with the most extra steps. As K decreases, these characters will receive progressively lower weights. For example, in Figure 3e, character 6 will receive a weight of 0.85 when K=6, but a weight of only 0.5 when K=1 (the "perfect" character 5, which has no extra steps, receives the maximum weight of 1). The optimal cladogram is that for which the summed weights for all characters has the largest value.

Consensus Cladograms (Trees)

A cladistic analysis will often produce more than one most parsimonious cladogram as a result of contradictory signal in the data (homoplasy). The agreement between

such “fundamental cladograms” (so-called because they are generated directly from the analysis of data) can be conveniently summarized by means of a consensus cladogram (usually referred to as a consensus tree). Several consensus methods have been proposed, of which the most widely used are “strict,” “ combinable components” (or “semistrict”), “Adams,” and “majority-rule.”
The strict consensus tree is the most conservative, because it includes only those groups (often referred to as “components”) that are common to all the fundamental cladograms. For example, in the two cladograms shown in Figures 7a and b, only groups ABC and DEF occur in both, and thus these are the only groups that appear in the strict consensus tree (Figure 7d). Groups EF, AB, and BC are excluded because the first is lacking from Figure 7b (where it is unresolved) and the other two are contradictory.

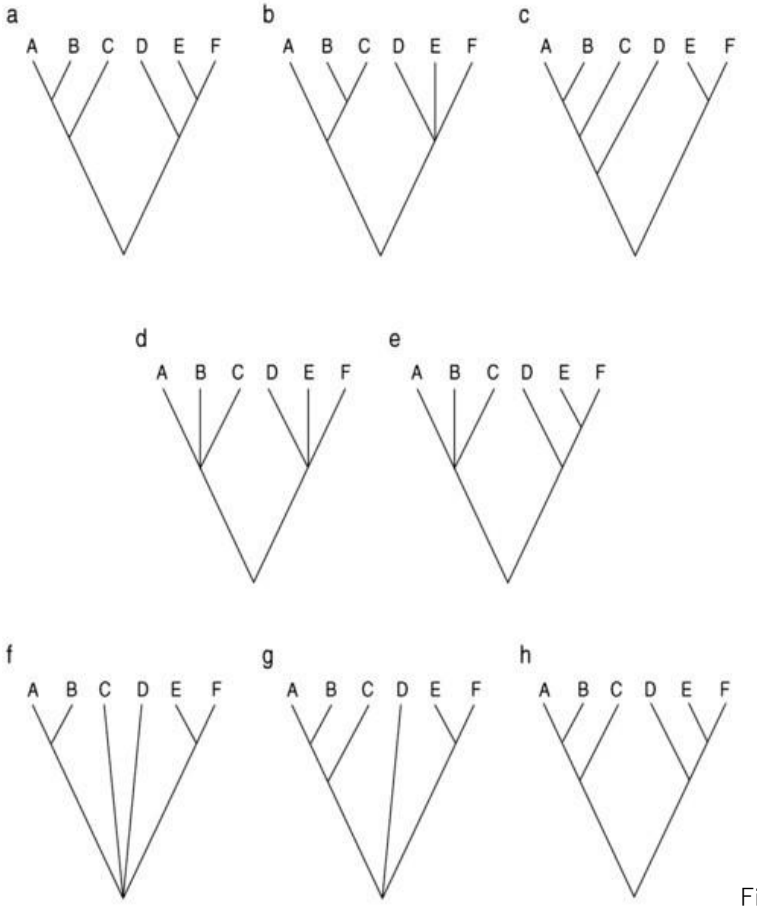

Figure 7. Consensus analysis. (a–c) Three cladograms for six taxa, A–F. (d) Strict consensus tree of cladograms 7a and 7b. (e) Combinable components consensus tree of cladograms 7a and 7b. (f) Strict consensus tree of cladograms 7a and 7c. (g) Adams consensus tree of cladograms 7a and 7c. (h) Majority-rule consensus tree of cladograms 7a, 7b, and 7c.

However, it is possible for a group to be lacking from one or more fundamental cladograms and yet be uncontradicted. For example, group EF in Figure 7a does not conflict with the cladogram in Figure 7b because it is one of the three resolutions

possible for the group DEF. Combinable components consensus allows such non-replicated, but non-conflicting, groups to be included in the consensus tree, in addition to those groups in common (Figure 7e). When all fundamental cladograms are fully resolved, with no spurious resolution due to ambiguous optimization, then the strict and combinable components consensus trees will be the same.

A problem with both strict and combinable components consensus is that a single taxon appearing in highly disparate positions on two cladograms is sufficient to collapse all intervening resolution. For example, taxon D in Figures 7a and c appears as the sister-group to two different terminal taxon-pairs. Consequently, the strict consensus tree (Figure 7f) is relatively unresolved. However, examination of Figures 7a and c shows that taxon D is acting as a "rogue" taxon; that is, apart from its differing positions, the resolution of the remaining taxa is identical in the two cladograms. Such rogue taxa can be identified using Adams consensus analysis. On an Adams consensus tree, taxa in conflicting positions on the fundamental cladograms are placed at the most inclusive node they have in common; in other words, the consensus contains all intersecting sets of taxa common to the fundamental cladograms. However, as a result, it is possible for an Adams consensus tree to contain groups that are not found on any of the fundamental cladograms and thus they need to be interpreted with care. Sometimes taxa such as D would simply be deleted to give the "largest common pruned tree" as a consensus.

When the number of fundamental cladograms is large, strict consensus trees are often very poorly resolved and can be viewed as too restrictive. One method of increasing resolution of the consensus is to retain those groups that occur in a prespecified number of the fundamental cladograms. Typically, such majority-rule consensus trees will comprise those groups that occur in more than 50% of cladograms. The majority-rule consensus tree of the cladograms in Figure 7a–c, shown in Figure 7h, is fully resolved, despite the marked topological differences in its fundamental cladograms.

Regardless of their number, the most parsimonious cladograms found by cladistic analysis remain our best estimate of the relationships among the taxa under study. Because resolution is lost, most consensus trees are less parsimonious than their

fundamental cladograms. However, if the topological differences among the fundamental cladograms are due solely to ambiguous optimization, the strict consensus tree will also be of minimum length. Then, the strict consensus tree is also the strictly supported cladogram (Nixon and Carpenter, 1996) and is the preferred most parsimonious topology because it is the only cladogram that is both of minimum length and has all groups supported unambiguously by data.

Cladogram and Group Support

A number of statistics attempt to assign levels of support or confidence to the results of cladistic analyses. They can be divided into two categories: methods that address support for an entire cladogram and aim to determine whether there is any “significant” structure in the data, and methods that examine the support afforded to individual groups on a cladogram and attempt to distinguish those groups that are well supported from those that are not.

Methods aimed at assessing support for an entire cladogram all use the same general principle. The length of the most parsimonious cladogram obtained from the observed data set is compared with those derived from a large number of “ phylogenetically uninformative” data sets, with the expectation that the former will be substantially shorter than any of the latter. Several definitions of “ phylogenetically uninformative” data have been proposed, including “statistically random” (scores in a data matrix are allocated at random), “undecisive” (all possible informative characters occur in equal numbers), and “randomly co-varying. ”

The simplest measure of support for a particular group on a cladogram is branch length. However, homoplasy makes the assessment of branch support difficult and groups may appear to be better supported than they actually are. “Bremer support” is a more precise measure of clade support and is defined as the number of extra steps required to lose a clade from the strict consensus tree of near-minimum-length cladograms. When there is no homoplasy in the data, the Bremer support of a group is the same as its branch length. Otherwise, support is reduced to the extent that there are alternative equally parsimonious groupings. To calculate Bremer support, first those cladograms that are one step longer than minimum are found and the strict consensus tree formed from them and the most parsimonious cladogram. The process is repeated, adding a step at a time, until the group in question is lost from the consensus. The number of extra steps required to achieve this is the Bremer support for the group. If more than one most parsimonious cladogram is found initially, then the procedure begins with the consensus tree of these cladograms. Any group that may be a potential resolution of the consensus will have a Bremer support equal to 0.

The bootstrap seeks to estimate group support using pseudoreplicate data sets, which are formed by randomly sampling characters with replacement. The effect is to weight some characters and delete others, with the constraint that the total weight equals the original number of characters. A large number of such pseudoreplicates are generated and their most parsimonious cladograms are found. Conflict among these cladograms is assessed using a majority-rule consensus tree and the support for a group is estimated as the proportion of most parsimonious cladograms on which it is recovered.

However, use of the bootstrap is questionable in a cladistic context.

The bootstrap assumes that a data set represents a random sample of all possible

The bootstrap assumes that a data set represents a random sample of all possible characters. However, taxonomic characters are generally carefully selected with the aim of resolving the relationships of the taxa under study. Furthermore, it only requires a single synapomorphy to diagnose a clade. However, the random nature of the bootstrap process means that such a character may be represented in only a

few pseudoreplicates and thus the group will not appear in the majority-rule consensus tree despite being uncontradicted. Thus, bootstrap values are a one-sided test; recovered groups have some measure of support in the data, but groups that are not recovered cannot be rejected.

An justifiable method for estimating support for groups on a cladogram, especially when characters are subjected to differential weighting, is “symmetric resampling” (Goloboff *et al.*, 2003), which calculates the difference between the absolute frequency with which a clade is found in resampled matrices and that in the most frequent alternative topology where the clade is not recovered (“Groups present / Contradicted” or GC values). GC values range from 100, where the clade is recovered in all resampled matrices, to -100, where an alternative arrangement is found in all resampled matrices, with a value of zero indicating that levels of support and contradiction are equal.

Simultaneous and Partitioned Analysis

It is generally recognized that data from many sources may be used in a cladistic analysis (e.g., morphological, physiological, behavioral, ecological, or molecular sequences) and that analyses of these data can yield different hypotheses of relationships. Simultaneous analysis (sometimes called a total evidence approach) combines all data, from whatever source, into a single taxon × character matrix for analysis. The resulting hypothesis of relationships is thus determined by character congruence (Figure 8a). Alternatively, different classes of data may be analyzed separately and the resulting cladograms added together using a consensus method to extract the common phylogenetic signal. This is called the partitioned evidence approach (Figure 8b) and the result is determined by taxic congruence. The reasoning underlying this approach is that different classes of data may reflect different evolutionary processes and so should be analyzed separately. A third alternative, known as conditional data combination, attempts to distinguish those conditions under which it would be best to keep data sets separate and conduct partitioned evidence analysis from those conditions under which it is more appropriate to conduct a simultaneous analysis. This approach estimates the

degree of heterogeneity of phylogenetic signal among data sets, and if the heterogeneity is greater than might be explained by sampling error, then the data sets are analyzed separately.

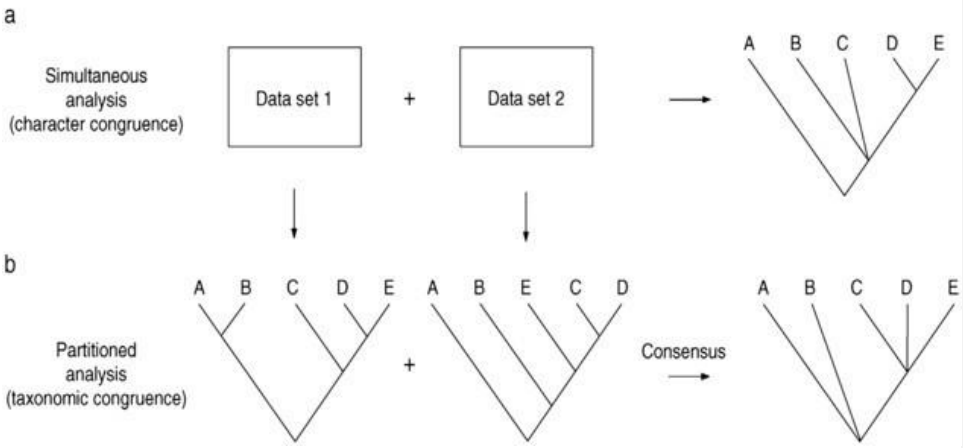

Figure 8. Simultaneous and partitioned analysis. In simultaneous analysis, all data are combined into a single matrix before analysis. In partitioned analysis, each data set is analyzed separately and the resulting cladograms are then “added” together using a consensus method.

See also

Cladogenesis. Diversity, Taxonomic versus Functional. Evolution, Theory of. Systematics, Overview. Taxonomy, Methods of

References

Farris, 1989. J.S. Farris . The retention index and the rescaled consistency index Cladistics 5 1989 417-419

Goloboff, 1993. P.A. Goloboff . Estimating character weights during tree search Cladistics 9 1993 83-91

Goloboff et al., 2003. P. Goloboff , J. Farris , M. Källersjö , B. Oxelmann , M. Ramírez , C. Szumik . Improvements to resampling measures of group support Cladistics 19 324–332.

Hennig, 1950. W. Hennig . Grundzüge einer Theorie der phylogenetischen Systematik 1950 Deutsche Zentralverlag Berlin

Hennig, 1966. W. Hennig . Phylogenetic Systematics 1966 University of Illinois Press Urbana

Kitching et al., 1998. I.J. Kitching , P.L. Forey , C.J. Humphries , D.M. Williams . Cladistics: The Theory and Practice of Parsimony Analysis 2nd ed 1998 Oxford University Press Systematics Association Publication 11, Oxford, United Kingdom

Maddison et al., 1984. W.P. Maddison , M.J. Donoaghue , D.R. Maddison . Outgroup analysis and parsimony Systematic Zool. 33 1984 83-103

Nelson, 1973. G.J. Nelson . The higher-level phylogeny of the vertebrates Systematic Zool. 22 1973 87-91

Nelson, 1978. G.J. Nelson . Ontogeny, phylogeny, paleontology, and the biogenetic law Systematic Zool. 27 1978 324-345

Nixon and Carpenter, 1996. K.C. Nixon , J.M. Carpenter . On consensus, collapsibility, and clade concordance Cladistics 12 1996 305-321

Schuh and Brower, 2009. R.T. Schuh , A.V.Z. Brower . Biological systematics: principles and application 2nd ed Comstock Publishing Associates, Cornell University Press.

Sereno, 2007. P.C. Sereno . Logical basis for morphological characters in phylogenetics Cladistics 23 565-587.

Watrous and Wheeler, 1981. L.E. Watrous , Q.D. Wheeler . The outgroup comparison method of character analysis Systematic Zool. 30 1981 1-11

Williams, 2004. D.M. Williams . Homology and homologues, cladistics and phenetics: 150 years of progress Pp. 191-224, in: Williams, D.M. & Forey, P.L., editors, *Milestones in Systematics* London: Taylor & Francis.

Change History: September 2015. IJ Kitching and DM Williams introduced small edits in the text of the article including citations, and made larger changes to the sections “Cladogram and Group Support” and “References”.